isolated repeats show < 10% overlap with our current set), we have reached the point of diminishing returns. The map covers the entire mouse genome, with the markers being sufficiently abundant, polymorphic and stable to allow the mapping of monogenic or polygenic traits in virtually any mouse cross of interest[3,8]. Moreover, the markers are sufficiently dense to facilitate positional cloning of most mouse mutations. With > 90% of the mouse genome being within 750 kb of a marker, and current mouse yeast artificial chromosome (YAC) libraries[14,15] having a mean insert size > 750 kb, the map affords ready access to the vast majority of the genome with little need for chromosomal walking, and provides a preliminary scaffold for constructing a genome-wide physical map[16].

The map also provides a common framework for the mapping of mutations and cloned genes. In addition to our integration with the Frederick cross, the SSLP map is being used as a framework for other mapping crosses, including public resources at the Jackson Laboratory[17] and the European Collaborative Interspecific Backcross (EUCIB)[18]. The EUCIB Project: the construction of dense genetic maps of mouse and man. □

Received 23 October 1995; accepted 19 February 1996.

1. Copeland, N. G. *et al. Science* **262**, 57–66 (1993).
2. Copeland, N.G. *et al. Science* **262**, 67–82 (1993).
3. Dietrich, W. F. *et al. Genetics* **131**, 423–447 (1992).
4. Dietrich, W. F. *et al.* in *Genetic Maps* 1992 (ed. O'Brien, S.) 4.110–4.142 (Cold Spring Harbor Laboratory Press, NY, 1992).
5. Miller, J. C. *et al.* in *Genetic Variants and Strains of the Laboratory Mouse* 3rd edn. (eds Lyons, M. F. & Searle, A.) (Oxford Univ. Press, New York, 1994).
6. Dietrich, W. F. *et al. Nature Genet.* **7**, 220–245 (1994).
7. Lander, E. S. *et al. Genomics* **1**, 174–181 (1987).
8. Lincoln, S. E. & Lander, E. S. *Genomics* **14**, 604–610 (1992).
9. Copeland, N. G. & Jenkins, N. A. *Trends Genet.* **7**, 113 (1991).
10. Ceci, J. D. *et al. Genomics* **8**, 699–709 (1989).
11. Buchberg, A. M. *et al. Genetics* **122**, 153–161 (1989).
12. Jacob, H. J. *et al. Nature Genet.* **9**, 63–69 (1995).
13. Dib, C. *et al. Nature* **380**, 152–154 (1996).
14. Lann, Z., Monaco, A. P. & Lehrach, H. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4123 (1991).
15. Kusumi, K. *et al. Mamm. Genome* **4**, 391–392 (1993).
16. Hudson, T. *et al. Science* **270**, 1945–1955 (1995).
17. Rowe, L. B. *et al. Mamm. Genome* **5**, 253–274 (1994).
18. The European Backcross Collaborative Group. *Hum. molec. Genet.* **3**, 621–627 (1994).
19. Evans, E. *in Genetic Variants and Strains of the Laboratory Mouse* 3rd edn. (eds Lyons, M. F. &





# A comprehensive genetic map of the human genome based on 5,264 microsatellites

**Colette Dib*, Sabine Fauré*, Cécile Fizames*, Delphine Samson*, Nathalie Drouot*, Alain Vignal*, Philippe Millasseau*, Sophie Marc*, Jamilé Hazan*, Eric Seboun*, Mark Lathrop†, Gabor Gyapay*, Jean Morissette*‡ & Jean Weissenbach*§**

* Généthon and CNRS URA 1922, 1 rue de l'Internationale, 91000 Evry, France
† INSERM U358, Hôpital Saint-Louis, Paris, France
‡ Centre de Recherche du Centre Hospitalier de l'Université Laval, Québec G1V 4G2, Canada
§ To whom correspondence should be addressed

**The great increase in successful linkage studies in a number of higher eukaryotes during recent years has essentially resulted from major improvements in reference genetic linkage maps[1–4], which at present consist of short tandem repeat polymorphisms of simple sequences or microsatellites[7,8]. We report here the last version of the Généthon human linkage map[6]. This map consists of 5,264 short tandem (AC/TG)$_n$ repeat polymorphisms with a mean heterozygosity of 70%. The map spans a sex-averaged genetic distance of 3,699 cM and comprises 2,335 positions, of which 2,032 could be ordered with an odds ratio of at least 1,000 : 1 against alternative orders. The average interval size is 1.6 cM; 59% of the map is covered by intervals of 2 cM at most and 1% remains in intervals above 10 cM.**

Microsatellite markers were obtained as described previously[5,6]. A heterozygosity above 0.5 was observed for 93% of the markers and above 0.7 for 58%. These values remain very close to those of our previous version[6]. Average heterozygosity per chromosome varied from 0.65 (chromosome X) to 0.73 (chromosome 19), with a mean value of 0.70 for the entire collection of markers (Table 1). Database sequence comparisons and searches detected matches of AFM (Association Française contre les Myopathies) markers with 19 genes and 74 anonymous markers.

Genotyping of the microsatellite markers was performed as described previously on the same eight CEPH (Centre d'Etudes du Polymorphisme Humaine) families (20 for the X chromosome), which comprised a total of 134 individuals and 186 meioses[5,6] (304 individuals and 291 meioses for the X chromosome). Genotypes were submitted to the same error-checking procedures as reported earlier[6]. These procedures consisted of (1) a reinvestigation of families with abnormally elevated recombination frequencies between pairs of markers, and (2) correction or elimination of all double recombinant genotypes of markers placed in short linkage intervals. Such apparent double recombinations probably result from mutation events that converted an allele of one individual into the other allele. A more detailed analysis of double-recombination events and mutations in microsatellites is in preparation.

Map construction was done in a stepwise manner with multiple controls at each step. The total length of this map as evaluated from the CILINK algorithm[9] is 3,699 cM (Table 1). This is almost identical in length to our previous version, despite the addition of new terminal markers that extend the 93/94 chromosome maps by 145 cM (4%). The absence of increase in length probably results from a very thorough error-checking process and from elimination of apparent double-recombinant genotypes. The 5,264 markers are distributed in 2,335 positions (Fig. 1), 2,032 of which are ordered with odds ratios against alternative orders of at least 1,000 : 1. The mean interval size is 1.6 cM. The fraction of the map in intervals above 10 cM represents only 1 per cent of the total linkage distance and consists of 3 intervals spanning 11 cM. Fifty-nine per cent of the map is covered by intervals of 2 cM at most, and 92 per cent by intervals of 5 cM at most. Markers from the CEPH and CHLC databases have been integrated into this map as shown in Fig. 2, which presents the map of chromosome 22 as an example. Detailed information, including integrated maps of all chromosomes, a list of markers, their primer sequences, heterozygosity, number and size-range of alleles observed in the 8 (or 20) genotyped CEPH families, sex-specific distances, and mutations, will be presented in an extended reprint available on request and on an electronic server (http://www.genethon.fr).

The total sex-specific lengths of autosomes estimated by CILINK[9] show only slight variations when compared to the lengths of the previous map[6]. The length excess observed for the female map is comparable to other published maps. This excess

TABLE 1 Quantitative characteristics of maps and markers by chromosome

| Chromosome | Physical length* (Mb) | Number of markers mapped | Markers per Mb | Genetic length (cM)† | | | Markers per cM | Mean heterozygosity‡ | Ordered 1,000 : 1§ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Female | Male | Sex-average | | | |
| 1 | 263 | 461 | 1.75 | 358.2 | 220.3 | 292.7 | 1.57 | 0.71 | 142 |
| 2 | 255 | 452 | 1.77 | 324.8 | 210.6 | 277.0 | 1.63 | 0.71 | 160 |
| 3 | 214 | 353 | 1.65 | 269.3 | 182.6 | 233.0 | 1.51 | 0.71 | 199 |
| 4 | 203 | 280 | 1.38 | 270.7 | 157.2 | 212.2 | 1.32 | 0.69 | 103 |
| 5 | 194 | 312 | 1.61 | 242.1 | 147.2 | 197.6 | 1.57 | 0.70 | 117 |
| 6 | 183 | 311 | 1.70 | 265.0 | 135.2 | 201.1 | 1.54 | 0.71 | 116 |
| 7 | 171 | 272 | 1.59 | 187.2 | 178.1 | 184.0 | 1.48 | 0.72 | 116 |
| 8 | 155 | 249 | 1.61 | 221.0 | 113.1 | 166.4 | 1.50 | 0.70 | 93 |
| 9 | 145 | 189 | 1.30 | 194.5 | 138.5 | 166.5 | 1.13 | 0.71 | 67 |
| 10 | 144 | 281 | 1.95 | 209.7 | 146.1 | 181.7 | 1.55 | 0.70 | 101 |
| 11 | 144 | 273 | 1.90 | 180.0 | 121.9 | 156.1 | 1.75 | 0.70 | 104 |
| 12 | 143 | 249 | 1.74 | 211.8 | 126.2 | 169.1 | 1.47 | 0.71 | 95 |
| 13q | 98 | 164 | 1.67 | 132.3 | 97.2 | 117.5 | 1.40 | 0.70 | 61 |
| 14q | 93 | 162 | 1.74 | 154.4 | 103.6 | 128.6 | 1.26 | 0.72 | 70 |
| 15q | 89 | 145 | 1.63 | 131.4 | 91.7 | 110.2 | 1.36 | 0.70 | 52 |
| 16 | 98 | 180 | 1.84 | 169.1 | 98.5 | 130.8 | 1.38 | 0.71 | 75 |
| 17 | 92 | 186 | 2.02 | 145.4 | 104.0 | 128.7 | 1.44 | 0.70 | 68 |
| 18 | 85 | 136 | 1.60 | 151.3 | 92.7 | 123.8 | 1.10 | 0.70 | 50 |
| 19 | 67 | 121 | 1.81 | 115.0 | 98.0 | 109.9 | 1.10 | 0.73 | 51 |
| 20 | 72 | 144 | 2.00 | 120.3 | 73.3 | 96.5 | 1.49 | 0.70 | 52 |
| 21q | 39 | 61 | 1.56 | 70.6 | 46.8 | 59.6 | 1.02 | 0.70 | 27 |
| 22q | 43 | 67 | 1.56 | 74.7 | 46.9 | 58.1 | 1.15 | 0.71 | 32 |
| X | 164 | 216 | 1.32 | 198.1 | | 198.1 | 1.09 | 0.65 | 81 |
| Genome | 3,154 | 5,264 | 1.67 | 4,396.9 | 2,729.7 | 3,699.2 | 1.38 | 0.70 | 2,032 |

* Physical lengths are from ref. 18. † Genetic lengths were determined using the CILINK program of the LINKAGE package[9]. ‡ Heterozygosities were determined using genotypes of 56 autosomes (or 42 X chromosomes) from unrelated caucasoid individuals (CEPH grandparents of families 884, 1331, 1332, 1347, 1362 and 1416 and parents of families 102 and 1413). § 1,000 : 1 odds order was determined as for Fig. 1.

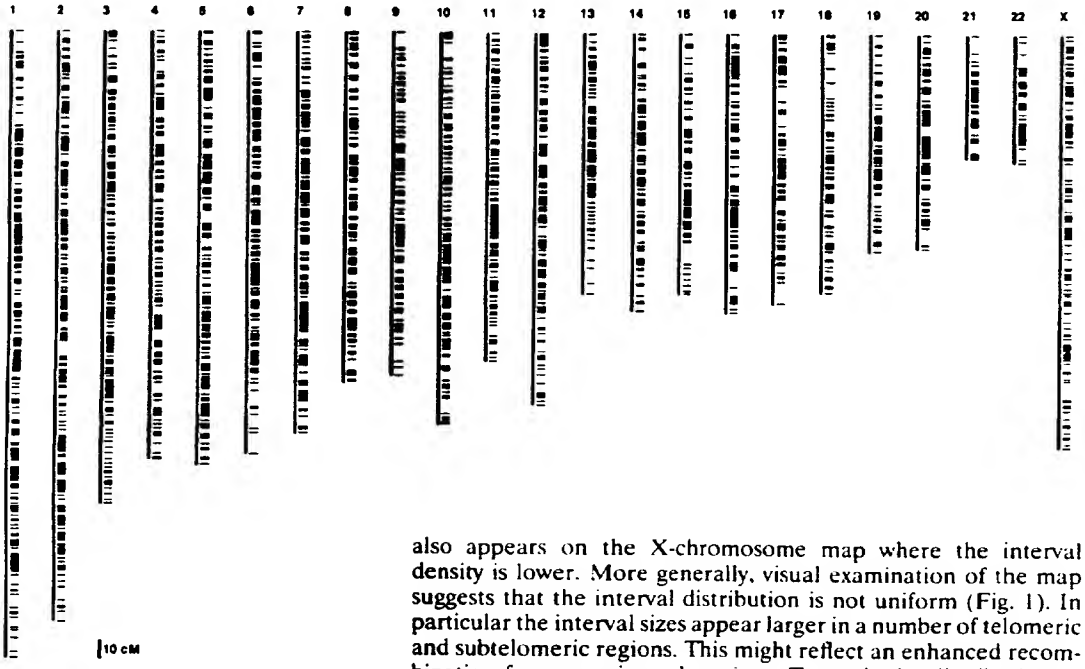


FIG. 1 Genetic linkage map of the 2,335 positions defined by 5,264 markers. The vertical bar delimits the length covered by the map as computed by the GMS algorithm[19] and represents sex-averaged genetic distances in cM. Horizontal marks represent map positions of AFM microsatellite markers. All distance values between positions were rounded to integer values and distances between 0 and 1 cM were rounded to 1 cM. Slight differences may appear between distances evaluated using GMS (this figure) and CILINK displayed in Table 1. This reflects the fact that clusters of non-recombinant markers were processed in a different manner by each program. Maps were constructed using the automated map construction algorithm MultiMap[20] based on CRIMAP[21]. The order proposed by MultiMap (primary map) was further submitted to another algorithm, GMS, based on the LINKAGE package[9]. For markers not ordered with the support of 1,000 : 1 odds on the primary map, the most likely order was chosen. Markers for which there were two equally probable and 'most likely' positions usually did not recombine with another marker situated between those two positions. Absence of recombination between such markers was verified and they were submitted as non-recombinants for the GMS computation. The order of loci proposed by the GMS algorithm (secondary map) was further checked for double recombinants. Given the high density of markers, these double recombinant genotypes were very unlikely and therefore subjected to rescoring. Among a total of 874 double recombinant genotypes rescored, 71% were reassessed. A second genotyping experiment was carried out for suspicious genotypes. This procedure was used until the latest secondary order no longer revealed new double recombinant genotypes. The GMS algorithm usually computes sets of 15 to 20 markers from segments of a chromosome. For the final GMS computation, these segments were defined manually to maximize the number of positions that could be ordered with 1,000 : 1 odds. A total of 128 double recombinants could not be corrected. These apparent double recombinations occurred in very small intervals surrounded by several markers of the alternative phase. Moving them by a few cM did not solve the problem or resulted in an increased number of alternative double recombinant genotypes on other haplotypes. Because such apparent double recombination events are most probably mutations, they were eliminated from the dataset for the final map calculations.

also appears on the X-chromosome map where the interval density is lower. More generally, visual examination of the map suggests that the interval distribution is not uniform (Fig. 1). In particular the interval sizes appear larger in a number of telomeric and subtelomeric regions. This might reflect an enhanced recombination frequency in such regions. To study the distribution of genetic markers throughout the human genome, we designed a simulation model that could account for variations resulting from marker identification and typing processes. On average, the observed numbers are close to the simulation using a density of two highly informative $(AC)_n$ microsatellites per cM without significant bias in the observed marker distribution (results not shown).

The physical sizes of the largest gaps were estimated by measuring the number of radiation-induced breaks in a panel of whole-genome radiation hybrids developed recently[10]. Assuming that breakage distances of radiation hybrid maps are representative of physical distances, 3 out of 4 large genetic intervals tested are notably smaller than expected from the linkage distance (results not shown). This suggests that a number of the gaps remaining on the genetic map correspond to regions with

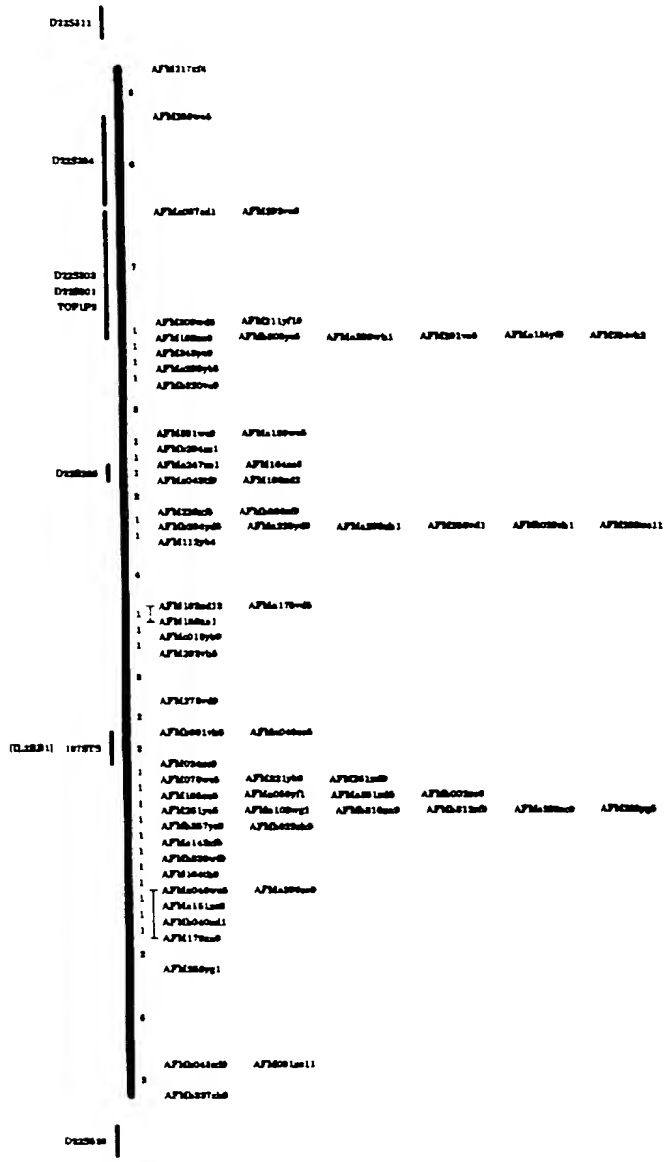

FIG. 2 Integrated genetic linkage map of chromosome 22. All chromosome maps are presented in the extended reprint, as illustrated in this example for chromosome 22. The leftmost column gives names of genes or non-AFM anonymous loci from the CEPH/CHLC databases. These loci were positioned by the MultiMap algorithm at 1,000 : 1 odds with respect to a framework of AFM markers. Gene names framed by brackets ([IL2RB1]) indicate genes for which a microsatellite was specifically developed as a genetic marker. These gene markers were genotyped, error-checked and positioned using the same criteria as for AFM markers. The vertical lines on the left indicate the intervals in which genes or non-AFM anonymous loci from the CEPH/CHLC databases could be placed by MultiMap at odds >1,000 : 1. The thick vertical bar delimits the length covered by the map as computed by the GMS algorithm. Some loci were placed at 1,000 : 1 odds outside the AFM framework; they are depicted on the top and/or bottom of the maps. Because genotypes of these loci were not submitted to the same quality control procedures as AFM markers, only relative positions, but not distances are presented, that is, their distance from other markers and the length of the interval corresponding to their position is arbitrary. AFM markers are displayed in the right part of the map from top to bottom according to the GMS defined order and distances. Markers appearing on the same line are markers with no recombinations in the informative subset of the 8 CEPH families tested. The order of AFM markers on a line is arbitrary. Numbers to the right of the thick bar between consecutive AFM markers correspond to sex-averaged genetic distances in cM evaluated by GMS between consecutive AFM markers. All values comprised between 0 and 1 cM were rounded to 1 cM. The vertical segments in front of the name of AFM markers indicate groups of markers for which the order could not be resolved with odds >1,000 : 1.

METHODS. Integration of other loci: sequence comparisons and searches in databases detected matches of AFM markers with 19 genes and 74 anonymous markers. These genes or loci, which had been independently assigned to the same chromosomes (except in one instance), could thus be precisely positioned on the present maps. We have attempted to include markers from the CEPH and CHLC databases in our maps. A framework map based on 1,844 AFM markers ordered at 1,000 : 1 odds was established using the MultiMap algorithm. An additional 186 non-AFM markers could be positioned at 1,000 : 1 odds on this predefined MultiMap framework and are indicated on the maps.

enhanced genetic recombination rather than to segments devoid of $(AC)_n$ microsatellites. However, as seen for one of the gaps analysed, some intervals may represent actual physical gaps. Similarly, clusters of non-recombining markers can be separated by distances estimated to be of the order of several megabases.

The highly informative microsatellites enabled us to order more than 2,000 markers with an odds ratio of 1,000 : 1 with the genotyping of only eight reference CEPH families representing a total of 186 informative meioses at best. But we gave priority to marker density rather than to degree of resolution. With this dense map and the additional genetic mapping resources available[2,11–15], linkage mapping of monogenic diseases can be readily carried out to the centimorgan level in most instances. In addition, the availability of numerous highly informative markers will also serve to define regions of linkage disequilibrium in which a founder haplotype can be reconstituted in genetically isolated populations. The genome-wide searches for loci involved in complex diseases will also considerably benefit from this new resource. In addition, this map provides the scaffold for the construction of physical maps based on overlapping sets of yeast artificial chromosomes that have been recently published[16,17]. □

Received 6 October 1995; accepted 8 January 1996.

1. Dietrich, W. F. *et al. Nature Genet.* **7**, 220–225 (1994).
2. Murray, J. C. *et al. Science* **265**, 2049–2070 (1994).
3. Barendse, W. *et al. Nature Genet.* **6**, 227–235 (1994).
4. Jacob, H. J. *et al. Nature Genet.* **9**, 63–69 (1995).
5. Weissenbach, J. *et al. Nature* **359**, 794–801 (1992).
6. Gyapay, G. *et al. Nature Genet.* **7**, 246–339 (1994).
7. Weber, J. L. & May, P. E. *Am. J. hum. Genet.* **44**, 388–396 (1989).
8. Litt, M. & Luty, J. A. *Am. J. hum. Genet.* **44**, 397–401 (1989).
9. Lathrop, G. M. & Lalouel, J. M. *Am. J. hum. Genet.* **36**, 460–465 (1984).
10. Walter, M. *et al. Nature Genet.* **7**, 22–28 (1994).
11. Sourn, N. K. *et al. Eur. J. hum. Genet.* **2**, 193–252 (1994).
12. Gerken, S. C. *et al. Am. J. hum. Genet.* **56**, 484–499 (1995).
13. Elsner, T. I., Albertsen, H., Gerken, S. C., Cartwright, P. & White, R. *Am. J. hum. Genet.* **56**, 500–507 (1995).
14. Plaetke, R. & Schachtel, G. *Am. J. hum. Genet.* **56**, 508–518 (1995).
15. Utah Marker Development Group *Am. J. hum. Genet.* **57**, 619–628 (1995).
16. Chumakov, I. M. *et al. Nature* **377**, 175–297 (1995).
17. Hudson, T. J. *et al. Science* **270**, 1945–1954 (1995).
18. Morton, N. E. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7474–7476 (1991).
19. Lathrop, G. M. *et al. Genomics* **2**, 157–164 (1988).
20. Matise, T. C., Perlin, M. & Chakravarti, A. *Nature Genet.* **6**, 384–390 (1994).
21. Lander, E. S. & Green, P. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2363–2369 (1987).

ACKNOWLEDGEMENTS. This work was initiated at CEPH and results from discussions with D. Cohen. It was supported by the Association Française contre les Myopathies, the Groupement de Recherches et d'Etudes sur les Génomes and European Union (Biomed1). We acknowledge the technical and clerical contribution of L. Baron, N. Becuwe, M. Besnard-Gonnet, I. Bordelais, C. Caboustan, C. Cruaud, M. Dubois, C. Dumont, C. Dupraz-Ramet, E. Ernst, K. Fonsat, M. François, J. L. Mangua, C. Marquette, M. Maugé, E. M. Bene, M. Meugnier, D. Muselet, S. Nguyen, S. Pezard, M. Tranchant, N. Sunn, N. Vega, E. Wunderle and V. Wunderle. The group of summer students substantially contributed to the data collection and error checking. We thank the informatics team of Généthon, particularly X. Benigni, L. Bougueleret, C. Discala, J.-M. Froussard, R. Gavrel, P. Gesnouin, S. Pook, P. Rodriguez-Tomé, L. Sainte-Marthe, C. Scarpelli and G. Vaysseix. We also thank A. de Sario and G. Bernardi for the DNA isochore fractionation, S. Cure for help in writing the manuscript, and G. Perrano for continuous devotion to this project.